



# OPEN DATA PUBLISHERS GUIDE

## Contents

Overview of the Connecticut Open Data Portal and the Socrata platform .....	2
Navigating a dataset .....	4
Preparing Your Data .....	7
Publishing a Dataset.....	12
LOCATION.....	15
METADATA.....	18

## Overview of the Connecticut Open Data Portal and the Socrata platform:

The dataset catalog is typically the first thing you see on a Socrata site, so what does everything mean on the catalog page?

The screenshot shows the Socrata 'Search & Browse Datasets and Views' page. On the left is a sidebar with filters, and on the right is a table of search results. Red numbers 1 through 8 are placed over specific UI elements to highlight key features.

**1.** Search bar at the top left.

**2.** View Types filter section, including Datasets, External Datasets, Files and Documents, Filtered Views, Charts, Maps, Calendars, and Forms.

**3.** Categories filter section, including Business, Education, Fun, Government, and Personal.

**4.** Topics filter section, including air, epa, filter, government, and water.

**5.** Search & Browse Datasets and Views header.

**6.** Sort dropdown menu set to 'Most Relevant'.

**7.** Table header with columns: Name, Popularity, and Type.

**8.** Pagination controls at the bottom, showing 'Showing 10 of 20075'.

	Name	Popularity	Type
1.	2010 Report to Congress on White House Staff	496,482 views	Government
2.	The White House - Nominations & Appointments	409,026 views	Government
3.	2011 Report to Congress on White House Staff	322,217 views	Government
4.	Milk RadNet Laboratory Analysis	209,281 views	Government
5.	Precipitation RadNet Laboratory Analysis	190,589 views	Government
6.	Le Sarkomètre	182,271 views	Government
7.	2012 Annual Report to Congress on White House Staff	155,663 views	Government
8.	Drinking Water RadNet Laboratory Analysis	146,553 views	Government
9.	Sorted RadNet Laboratory Analysis	144,341 views	Government
10.	Air Filter & Cartridge RadNet Laboratory Analysis	110,879 views	Government

**1. Search:** Enter in a keyword or words to search the catalog for matches. Results will be presented in the following order: matches in the title, matches in the dataset, and matches in the metadata. Standard boolean operators can be used, meaning you can type AND or OR between words to specify the type of search performed; an AND search is performed if no operator is specified. We also use word stemming, which means that if you type "educational," you will also get results that match education, educating, educate, etc.

**2. View Types:** Clicking on one of these will filter the catalog with data of only that type.

- **Datasets:** These are datasets uploaded into the Socrata platform.
- **External Datasets:** links to datasets stored elsewhere on the web.

- **Files and Documents:** downloadable files such as PDFs or Word documents
- **Filtered Views:** These are saved filtered Datasets. This is a view of a the type Dataset, and will be linked to the original Dataset through the "More Views" tab.
- **Charts:** Charts such as pie charts and bar charts created from Datasets. These will also be linked to the original Dataset through the "More Views" tab.
- **Maps:** Maps created from geolocation data or geospatial data (shapefile, kml, or kmz file). If it is based on geolocation data, it will also be linked to the original Dataset through the "More Views" tab.
- **Calendars:** Calendars created from Date & Time data in a Dataset. These will be linked to the original Dataset through the "More Views" tab.
- **Forms:** Forms are created from a Dataset, to collect information from users. These will be linked to the original Dataset through the "More Views" tab.

**3. Categories:** Clicking on a category will filter the catalog for all datasets with that category selected in the dataset metadata. These categories are set-up and customized by Administrators through the Admin panel.

**4. Topics:** Clicking on a Topic will filter the catalog by all datasets tagged with that word. This section is populated with terms as they are tagged on Datasets in the metadata.

**5. RSS Subscription:** Clicking on this allows users to subscribe to updates on the catalog.

**6. Catalog Display Format:** Toggle between detailed and rich display views of the catalog.

**7. Sort:** You can sort the catalog in a number of different ways:

- **Most Relevant:** This is generally associated with a search term, and is calculated based on a normalization of the number of views.
- **Most Accessed:** You can select "This Week," "This Month," and "This Year" and is based on the number of views of the page, or "Popularity."
- **Alphabetical:** Sorted alphanumerically.
- **Newest:** Sorted by the most recently added datasets.

- Oldest: Sorted by the least recently added datasets.
- Recently Updated: Sorted by the most recently updated datasets. This also takes into account newly added datasets, as their most recent update is the date of upload.
- Highest Rated: Sorted based on the average rating given to the dataset by users on the site.
- Most Comments: Sorted by the dataset with the most comments.

**8. Number of Datasets in the catalog and the page:** The text "Showing 10 of 20075" indicates the number of dataset on the page you are viewing, out of the total number of datasets and views in the catalog. For example, as you move through the pages, these numbers will not change.

## Navigating a dataset

Once you have found an interesting dataset, chart, or map to look at, what can you do on the page?

The screenshot shows a dataset page with the following elements:

- 1.** Title: "Nonprofit Corporations by Cities or Zip Codes"
- 2.** Subtitle: "Based on Active Nonprofit Corporations"
- 3.** Description: "Nonprofit corporations active on the record of the Secretary of State Corporation Division as of the first working day of the month."
- 4.** Social Media icons (RSS, Facebook, Twitter, Email)
- 5.** View options (Table, Map, Chart)
- 6.** Search bar: "Find in this Dataset"
- 7.** Action buttons (Manage, More Views, Filter, Visualize, Export, Discuss, Embed, About)

	Registry Number	Business Name	Entity Type	Regis
1	299818	UNITED METHODIST CHURCH, OREGON CITY	DOMESTIC NONPROFIT CORPORATION	17-M
2	299818	UNITED METHODIST CHURCH, OREGON CITY	DOMESTIC NONPROFIT CORPORATION	17-M
3	299818	UNITED METHODIST CHURCH, OREGON CITY	DOMESTIC NONPROFIT CORPORATION	17-M
4	299818	UNITED METHODIST CHURCH, OREGON CITY	DOMESTIC NONPROFIT CORPORATION	17-M
5	299818	UNITED METHODIST CHURCH, OREGON CITY	DOMESTIC NONPROFIT CORPORATION	17-M

Filter menu options: Conditional Formatting, Sort & Roll-Up, Filter (selected). Filter this dataset based on contents.

For reference, "View" means any filter, chart, map, etc that is "Saved" from a dataset.

**1. Title:** This is the title given to the dataset or view. Views, such as charts, can have a different name from the dataset it is based on.

**2. Based On:** If you are on a view, this will link to the dataset that the view is based on. For example, if you create a pie chart of the total number of governors for each party, this would link back to the dataset with the numbers for each year.

**3. Description:** This is the description given to the dataset or view.

**4. Social Media:** This allows you to interact with this dataset or view through social media:

- RSS Subscription: Subscribe to updates on the dataset

- Facebook: Share the dataset or view on Facebook
- Twitter: Share the dataset or view on Twitter
- Email: Share the dataset or view over email

**5. Dataset View:** These buttons allow you to switch between the different views of a dataset or view, you can have more than one view splitting the screen vertically at a time

- Detail Row: This view displays each column and row in a tabular format (seen in example above)
- Fat Row: This view displays the details for each row grouped together, for example columns may be in a vertical list.
- Single Row: This view displays all the details for one row at a time, There are arrows to scroll through next and previous rows.
- Chart: If you are viewing a chart a button for the chart view will appear.
- Map: If you are viewing a map, a button for the map view will appear.

**6. Search:** Enter in a word or words to [search within the dataset](#) for matches.

**7. Full Screen:** Make the dataset or view full screen on your display.

**8. Sidebar Buttons:** Each of these buttons opens up a sidebar on the right-hand side of the dataset or view. Not all will appear, depending on if you are logged in and what role you have on the site. For more about user roles, [read this article](#).

- Edit: Click this sidebar to create a [working copy](#) and edit the dataset. From a working copy, you can edit individual cells, [append new rows](#), or [replace the entire dataset](#).
- Manage: Click this sidebar to [transfer ownership](#) of the dataset or view to another user, delete the dataset or view, [share the dataset](#) or view with other users, [make the dataset public or private](#), show and hide columns, and change the column order.
- More Views: Click this sidebar to see the other views created from the same dataset. These might be filtered views, charts, or maps. The number in red reflects the total number of views created from the dataset.

- Filter: Click this sidebar to [sort the dataset](#) by columns, [group and roll up](#) the dataset, [filter the dataset](#), and set the default filter for that view.
- Visualize: Click this sidebar to set [conditional formatting](#), create a calendar, [create a map](#) (if you have location data), and [create a chart](#).
- Export: Click this sidebar to access API information, print the dataset, and [export the dataset](#).
- Discuss: If commenting is enabled, click this sidebar to [comment](#) and read other's comments on the dataset or view. Commenting can be enabled for the dataset and cell level.
- Embed: Click this sidebar to [create a form](#) from the dataset and create Social Data Player embeds
- About: Click this sidebar to view the metadata information about the dataset, edit the metadata, view dataset analytics, and [contact the dataset owner](#).

## Preparing Your Data

The Connecticut Open Data Portal provides a way to take data from spreadsheets and turn it into an online interactive dataset for public consumption over the internet. In essence, it delivers many of the same features available through desktop software; such as search, sort, and filter. In order to make the publishing process as simple and straightforward as possible; as well as to take full advantage of these functions, we sometimes need to change the way we store data in spreadsheets. Programs like Microsoft Excel give users significant flexibility in the way we provide information to the public, however spreadsheets can also be formatted in a way, or include additional elements that are not actually part of a dataset.

Unsaved View | Save As... | Revert  
Based on CT DCF Differential Response System Reports Accepted by Response Time and Type

	DATA_AS_OF	SFY	REGION	OFFICE	RESPONSE_TIME	TOTAL_RESPONSES	INTAKE_RE
1	09/15/2014	SFY2010	OTHER	Special Invest. Unit Office	24 Hours	115	
2	09/15/2014	SFY2010	OTHER	Special Invest. Unit Office	72 Hours	390	
3	09/15/2014	SFY2010	OTHER	Special Invest. Unit Office	Same Day	40	
4	09/15/2014	SFY2010	OTHER	Special Invest. Unit Office	Unassigned	1	
5	09/15/2014	SFY2011	OTHER	Special Invest. Unit Office	24 Hours	149	
6	09/15/2014	SFY2011	OTHER	Special Invest. Unit Office	72 Hours	512	
7	09/15/2014	SFY2011	OTHER	Special Invest. Unit Office	Same Day	41	
8	09/15/2014	SFY2012	OTHER	Special Invest. Unit Office	24 Hours	147	
9	09/15/2014	SFY2012	OTHER	Special Invest. Unit Office	72 Hours	444	
10	09/15/2014	SFY2012	OTHER	Special Invest. Unit Office	Same Day	36	
11	09/15/2014	SFY2013	OTHER	Special Invest. Unit Office	24 Hours	116	
12	09/15/2014	SFY2013	OTHER	Special Invest. Unit Office	72 Hours	388	

Filter sidebar: REGION is OTHER

The following will provide some examples of spreadsheets that are typically formatted in manner that, while still providing useful information, often make data difficult to not only import to the Open Data Portal also can limit the utility of the data in general.

**\*\*Please note the examples use real publicly available datasets. These examples are not intended to be critical of the Agencies that publish these resources.**

### 1.) Including descriptive information or “Metadata” in the spreadsheet.

The example below contains very descriptive information within the spreadsheet. This information is critical for end users to understand what data is contained within the spreadsheet, what each column represents, and how values in those columns were derived. The Open Data Portal will allow for Metadata and deceptive elements to be incorporated, however it needs to be removed from the sheet itself prior to import:



Copy of aen2014 [Compatibility Mode] - Excel

2013-14 Adjusted Equalized Net Grand List per Capita (AENGLC) and AENGLC Rank for the 169 Towns and Member Population Weighted AENGLC Ranks for the Regional School Districts and Regional Educational Service Centers					
		Item A. Highest 2010 Per Capita Income (PCI):		\$100,824	
District Code	District Name	(1) Equalized Net Grand List (ENGL) 2009	(2) Population 2010	(3) Per Capita Income (PCI) 2010	(4) AENGLC Rank / Weighted AENGLC Rank (Col 1 / Col 2) x (Col 3 / Item A)
001	ANDOVER	377,034,536	3,303	38,710	43,826.00
002	ANSONIA	1,484,130,265	19,249	26,225	20,054.67
003	ASHFORD	454,127,481	4,317	32,842	34,265.84
004	AVON	3,649,823,882	18,098	55,879	111,770.20
005	BARKHAMSTED	533,907,239	3,799	34,775	48,472.98
006	BEACON FALLS	711,752,725	6,049	32,710	38,173.52
007	BERLIN	3,269,406,578	19,866	38,134	62,245.35
008	BETHANY	864,122,513	5,563	47,241	72,781.57
009	BETHEL	2,815,952,632	18,584	36,608	55,017.17

AENGLC09

By removing the descriptive information, and providing more detailed column headers in the example below, we are now left with simple rows and columns. This will easily import into the Open Data Portal, and allow users to access a “clean” dataset.

Copy of aen2014 [Compatibility Mode] - Excel

A	B	C	D	E	F	G	H
District Code	District Name	Equalized Net Grand List 2009	Population 2010	Per Capita Income	Adjusted Equalized Net Grand List	AENGLC RANK	
001	ANDOVER	377,034,536	3,303	38,710	43,826.00	107	
002	ANSONIA	1,484,130,265	19,249	26,225	20,054.67	161	
003	ASHFORD	454,127,481	4,317	32,842	34,265.84	130	
004	AVON	3,649,823,882	18,098	55,879	111,770.20	26	
005	BARKHAMSTED	533,907,239	3,799	34,775	48,472.98	93	
006	BEACON FALLS	711,752,725	6,049	32,710	38,173.52	121	
007	BERLIN	3,269,406,578	19,866	38,134	62,245.35	60	
008	BETHANY	864,122,513	5,563	47,241	72,781.57	47	
009	BETHEL	2,815,952,632	18,584	36,608	55,017.17	77	
010	BETHLEHEM	662,933,592	3,607	39,704	61,458.42	63	
011	BLOOMFIELD	2,963,847,920	20,486	39,738	57,021.79	72	
012	BOLTON	660,466,933	4,980	42,312	55,657.20	75	
013	BOZRAH	362,193,839	2,627	38,339	52,427.34	84	
014	BRANFORD	4,939,663,243	28,026	41,540	72,617.08	48	
015	BRIDGEPORT	9,790,215,961	144,229	19,854	13,368.69	165	
016	BRIDGEWATER	671,263,810	1,727	58,172	190,851.00	13	
017	BRISTOL	6,111,926,027	60,477	29,629	29,698.94	140	
018	BROOKFIELD	3,315,269,259	16,452	49,705	99,342.77	32	
019	BROOKLYN	739,271,794	8,210	25,124	22,438.09	156	
020	BURLINGTON	1,265,793,121	9,301	43,392	58,570.49	67	
021	CANAAN	263,485,984	1,234	37,283	78,956.75	40	
022	CANTERBURY	543,548,196	5,132	30,453	31,990.25	135	
023	CANTON	1,572,126,580	10,292	46,401	70,299.32	51	
024	CHAPLIN	226,642,233	2,305	32,188	31,390.63	136	
025	CHESHIRE	4,133,444,441	29,261	40,498	56,740.43	74	
026	CHESTER	651,245,975	3,994	40,783	65,955.69	54	
027	CLINTON	2,222,717,758	13,260	37,117	61,709.17	61	

AENGLC09

Look Out Below! – Many spreadsheets also contain footnotes, disclaimers, or other information at the end of the rows of data. This is also vital information to end users, but it will be loaded with the dataset itself, thus affecting the usability. The Open Data Portal provides for this to be associated with the dataset, however it should be removed prior to importing.

Copy of aen2014 [Compatibility Mode] - Excel

	A	B	C	D	E	F	G
190	243	CES					58
191	244	ACES					128
192	245	LEARN					85
193	253	Eastconn					133
194	901	Norwich Free Acad.					143
195	902	Gilbert School					139
196	903	...					127
197							
198		Total	537,174,836.838	3,562,126	6,737,447	13,273,705.90	
199							
200		Data Sources:					
201							
202		ENGL 2009	Office of Policy and Management				
203		PCI 2010	U.S. Bureau of the Census (Collected through American Community Survey)				
204		Population 2010	U.S. Bureau of the Census (Collected through American Community Survey)				
205							

AEENGLC09

https://data.ct.gov/Health-and-Human-Services/CT-DCF-Differential-Response-System-Reports-Accepted-by-Response-Time-and-Type

CONNECTICUT OPEN DATA

CT DCF Differential Response System Reports Accepted by Response Time and Type

This dataset contains aggregate data by State Fiscal Year for all reports of abuse/neglect accepted by DCF for either a traditional Child Protective Services (CPS) investigation, or as of SFY2012, a Family Assessment Response (FAR). Figures are provided by mandated Response Time and Response Type, for each DCF Area Office each SFY beginning with 2005. Each report accepted is screened for safety and risk factors, and assigned an amount of time within which the agency is required to respond to the report. Mandated response times include "Same Day", "24 Hours", and "72 Hours". Traditionally, DCF has had only one manner of responding to such reports, which was a mandated Child Protective Services (CPS) investigation. As of April 2012, DCF began responding to low-risk reports through a voluntary Family Assessment Response (FAR) process. Reports handled through a FAR response still contain allegations that meet the statutory definitions of neglect, but they do not require a decision concerning whether they are substantiated or not. This policy has resulted in fewer substantiated allegations since its implementation, but the agency continues to serve as many or more families reported for abuse/neglect.

	DATA_AS_OF	SFY	REGION	OFFICE	RESPONSE_TIME	TOTAL_RESPONSES	INTAKE...
1	09/15/2014	SFY2010	OTHER	NULL	24 Hours	13	
2	09/15/2014	SFY2010	OTHER	NULL	72 Hours	43	
3	09/15/2014	SFY2010	OTHER	NULL	Same Day	1	
4	09/15/2014	SFY2010	OTHER	Central Office	24 Hours	45	
5	09/15/2014	SFY2010	OTHER	Central Office	72 Hours	80	
6	09/15/2014	SFY2010	OTHER	Central Office	Same Day	5	
7	09/15/2014	SFY2010	OTHER	Special Invest. Unit Office	24 Hours	115	
8	09/15/2014	SFY2010	OTHER	Special Invest. Unit Office	72 Hours	390	
9	09/15/2014	SFY2010	OTHER	Special Invest. Unit Office	Same Day	40	
10	09/15/2014	SFY2010	OTHER	Special Invest. Unit Office	Unassigned	1	

Suggest a Dataset Terms of Service Privacy Contact Us

© 2014 State of Connecticut Powered by Socrata

## 2.) Excessive Whitespace and Formatting:

The following example displays a spreadsheet, with very useful information. This spreadsheet is likely organized and formatted in a way that is intended to be printed by the end user. In essence, this sheet has empty cells, or “whitespace”, intended to provide a more visually appealing product. Unfortunately this formatting limits the utility of the data contained in the spreadsheet to users who may want to interact with the data. In the example, there are four measures (LF, EMP, UN, and RT) present for each town. Thus, if a user wanted to sort the sheet to find the town with the highest EMP value for May, the sheet would quickly become disorganized. In general, every cell within a spreadsheet should contain a value.

LAUS Substate 2013 R [Read-Only] [Compatibility Mode] - Excel

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW DEVELOPER Fuzzy Lookup POWERPivot TEAM

Clipboard Font Alignment Number

2013 (2014 Benchmark)  
CONNECTICUT LABOR FORCE DATA BY PLACE OF RESIDENCE  
TOWNS

Phone (860) 263-6293

Whitespace Metadata Whitespace

		JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	AY	AGE
Andover	LF	1,895	1,884	1,884	1,879	1,888	1,905	1,927	1,907	1,891	1,886	1,896	1,890	1,895	
	EMP	1,750	1,753	1,758	1,762	1,773	1,783	1,792	1,778	1,778	1,778	1,786	1,781	1,773	
	UN	145	131	126	117	115	122	135	129	113	108	110	109	122	
	RT	7.7	7.0	6.7	6.2	6.1	6.4	7.0	6.8	6.0	5.7	5.8	5.8	6.4	
Ansonia	LF	9,438	9,403	9,399	9,443	9,535	9,655	9,802	9,691	9,513	9,469	9,455	9,376	9,515	
	EMP	8,326	8,295	8,354	8,388	8,487	8,584	8,696	8,630	8,512	8,512	8,541	8,523	8,487	
	UN	1,112	1,108	1,045	1,055	1,048	1,071	1,106	1,061	1,001	957	914	853	1,028	
	RT	11.8	11.8	11.1	11.2	11.0	11.1	11.3	10.9	10.5	10.1	9.7	9.1	10.8	
Ashford	LF	2,501	2,490	2,490	2,484	2,508	2,551	2,563	2,531	2,480	2,486	2,494	2,483	2,505	
	EMP	2,291	2,301	2,306	2,317	2,347	2,357	2,376	2,344	2,328	2,333	2,347	2,347	2,333	
	UN	210	189	184	167	161	194	187	187	152	153	147	136	172	
	RT	8.4	7.6	7.4	6.7	6.4	7.6	7.3	7.4	6.1	6.2	5.9	5.5	6.9	
Avon	LF	8,980	8,983	8,989	8,974	9,070	9,136	9,177	9,110	9,098	9,100	9,097	9,048	9,064	
	EMP	8,514	8,520	8,546	8,557	8,599	8,654	8,684	8,632	8,640	8,648	8,681	8,654	8,611	
	UN	466	463	443	417	471	482	493	478	458	452	416	394	453	
	RT	5.2	5.2	4.9	4.6	5.2	5.3	5.4	5.2	5.0	5.0	4.6	4.4	5.0	
Barkhamsted	LF	2,286	2,290	2,288	2,274	2,297	2,300	2,315	2,286	2,264	2,274	2,276	2,269	2,285	
	EMP	2,103	2,110	2,120	2,126	2,147	2,156	2,166	2,150	2,141	2,145	2,150	2,145	2,138	

LAUS File Towns LMA's Bridgeport LMA Danbury LMA Enfield LMA ...

Fortunately, this same Excel file, contains a worksheet with “raw data” that will import quite nicely into the Open Data Portal. The example below contains basically the same data that would allow an end user to derive significantly more value.

LAUS Substate 2013 R [Read-Only] [Compatibility Mode] - Excel

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW DEVELOPER Fuzzy Lookup POWERPivot TEAM

Clipboard Font Alignment Number

General

AutoSum

Fill

Sort & Find

Filter

Clear

Editing

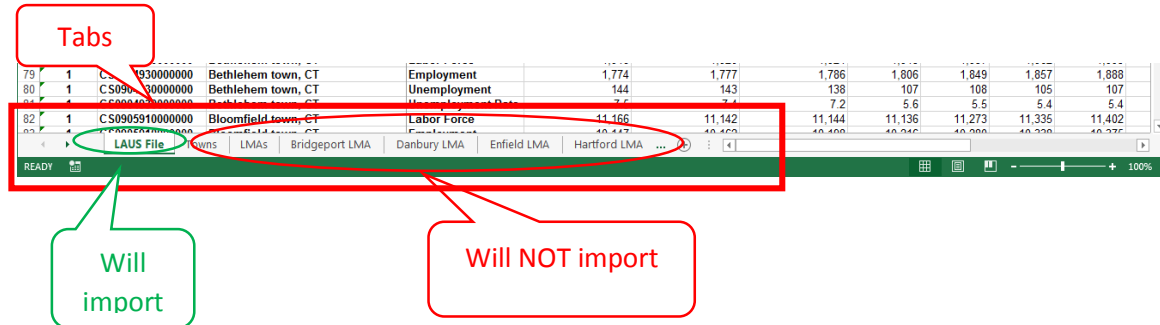
D10 Labor Force

	A	B	C	D	E	F	G	H	I	J	K
	Method	Area Code	Area Title	Data Type	Jan 2013 (B - 2014)	Feb 2013 (B - 2014)	Mar 2013 (B - 2014)	Apr 2013 (B - 2014)	May 2013 (B - 2014)	Jun 2013 (B - 2014)	Jul 2013 (B - 2014)
53	1	CS0901430000000	Ashford town, CT	Unemployment Rate	8.4	7.6	7.4	6.7	6.4	7.6	7.3
54	1	CS0902060000000	Avon town, CT	Labor Force	8,980	8,983	8,989	8,974	9,070	9,136	9,177
55	1	CS0902060000000	Avon town, CT	Employment	8,514	8,520	8,546	8,557	8,599	8,654	8,684
56	1	CS0902060000000	Avon town, CT	Unemployment	466	463	443	417	471	482	493
57	1	CS0902060000000	Avon town, CT	Unemployment Rate	5.2	5.2	4.9	4.6	5.2	5.3	5.4
58	1	CS0902760000000	Barkhamsted town, CT	Labor Force	2,286	2,290	2,288	2,274	2,297	2,300	2,315
59	1	CS0902760000000	Barkhamsted town, CT	Employment	2,103	2,110	2,120	2,126	2,147	2,156	2,166
60	1	CS0902760000000	Barkhamsted town, CT	Unemployment	183	180	168	148	150	144	149
61	1	CS0902760000000	Barkhamsted town, CT	Unemployment Rate	8.0	7.9	7.3	6.5	6.5	6.3	6.4
62	1	CS0903250000000	Beacon Falls town, CT	Labor Force	3,325	3,311	3,318	3,329	3,363	3,395	3,441
63	1	CS0903250000000	Beacon Falls town, CT	Employment	3,085	3,077	3,102	3,110	3,148	3,173	3,193
64	1	CS0903250000000	Beacon Falls town, CT	Unemployment	240	234	216	211	215	222	248
65	1	CS0903250000000	Beacon Falls town, CT	Unemployment Rate	7.2	7.1	6.5	6.3	6.4	6.5	7.2
66	1	CS0904300000000	Berlin town, CT	Labor Force	11,349	11,367	11,360	11,342	11,447	11,499	11,572
67	1	CS0904300000000	Berlin town, CT	Employment	10,572	10,598	10,636	10,662	10,750	10,803	10,851
68	1	CS0904300000000	Berlin town, CT	Unemployment	777	769	724	680	697	696	721
69	1	CS0904300000000	Berlin town, CT	Unemployment Rate	6.8	6.8	6.4	6.0	6.1	6.1	6.2
70	1	CS0904580000000	Bethany town, CT	Labor Force	2,987	2,998	2,991	3,008	3,052	3,073	3,079
71	1	CS0904580000000	Bethany town, CT	Employment	2,804	2,807	2,817	2,843	2,867	2,887	2,890
72	1	CS0904580000000	Bethany town, CT	Unemployment	183	191	174	165	185	186	189
73	1	CS0904580000000	Bethany town, CT	Unemployment Rate	6.1	6.4	5.8	5.5	6.1	6.1	6.1
74	1	CS0904720000000	Bethel town, CT	Labor Force	10,335	10,301	10,361	10,328	10,491	10,574	10,713
75	1	CS0904720000000	Bethel town, CT	Employment	9,922	9,956	9,965	9,980	9,812	9,871	9,972
76	1	CS0904720000000	Bethel town, CT	Unemployment	713	715	706	648	679	703	741
77	1	CS0904720000000	Bethel town, CT	Unemployment Rate	6.9	6.9	6.8	6.3	6.5	6.6	6.9
78	1	CS0904930000000	Bethlehem town, CT	Labor Force	1,918	1,920	1,924	1,913	1,957	1,962	1,995
79	1	CS0904930000000	Bethlehem town, CT	Employment	1,774	1,777	1,786	1,806	1,849	1,857	1,888
80	1	CS0904930000000	Bethlehem town, CT	Unemployment	144	143	138	107	108	105	107
81	1	CS0904930000000	Bethlehem town, CT	Unemployment Rate	7.5	7.4	7.2	5.6	5.5	5.4	5.4
82	1	CS0905910000000	Bloomfield town, CT	Labor Force	11,166	11,142	11,144	11,136	11,273	11,335	11,402

LAUS File Towns LMA's Bridgeport LMA Danbury LMA Enfield LMA Hartford LMA ...

### 3.) Multiple Tab Workbooks:

Often times Excel workbooks are broken up into multiple tabs. In some instances this may be to provide “lookup” values when one wishes to standardize or predefine values that are eligible during the data entry process. In other instances, users do this to break up larger datasets or to present different views of data. In general, data has the most value to users when it’s not broken up into subsets. More importantly however, the Open Data Portal, will only import the first tab in any given workbook. Thus, it is critical to ensure that the most robust set of data is compiled into a single sheet or tab within Excel, and that that specific tab is the first one in the sheet.



### 4.) “RAW DATA”

Ideally, data should be posted in its rawest form. This means the least amount of aggregation possible. In some instances, it is necessary to aggregate data, so that it cannot be used to identify an individual. However, this is only applicable when such data is protected from disclosure by State or Federal law.



## Getting to "raw" fruits and vegetables

Least flexible		
Category	Type	Purchased
Fruit	Strawberries	42
Fruit	Banana	36
Fruit	Oranges	101
Vegetable	Kale	52
Vegetable	Arugula	29
Vegetable	Green beans	67

More flexible			
Category	Type	Purchased	Year
Fruit	Strawberries	22	2014
Fruit	Strawberries	20	2013
Fruit	Banana	16	2014
Fruit	Banana	20	2013
Fruit	Oranges	61	2014
Fruit	Oranges	40	2013
Vegetable	Kale	30	2014
Vegetable	Kale	22	2013
Vegetable	Arugula	15	2014
Vegetable	Arugula	14	2013
Vegetable	Green beans	35	2014
Vegetable	Green beans	32	2013

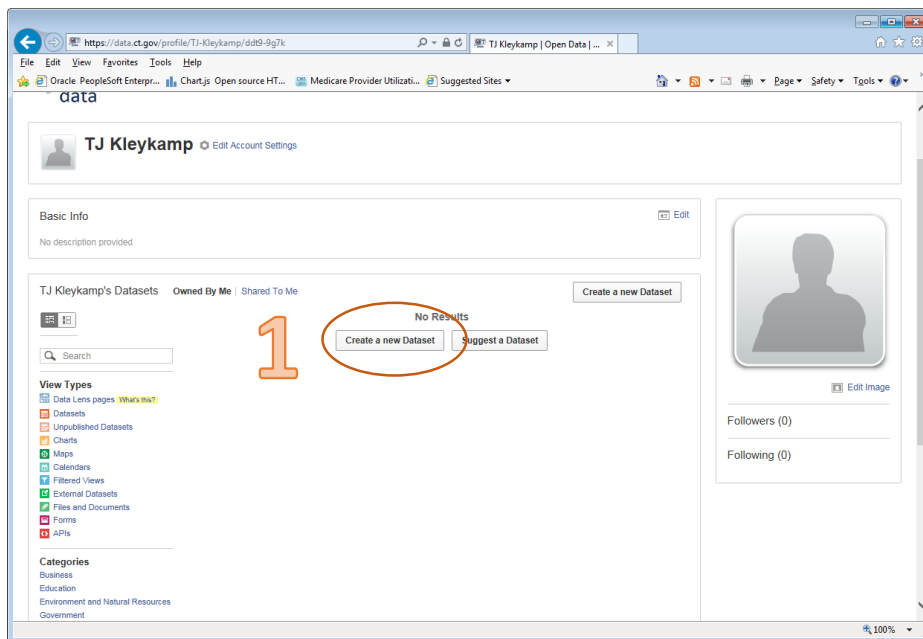
Raw! Preferred			
Unique ID	Category	Type	Purchase date
25545	Fruit	Strawberries	4/5/2014
45229	Fruit	Strawberries	4/6/2014
41545	Fruit	Strawberries	4/7/2014
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.

Eat your  
veggies!

(Image source: datasf.org)

## Publishing a Dataset

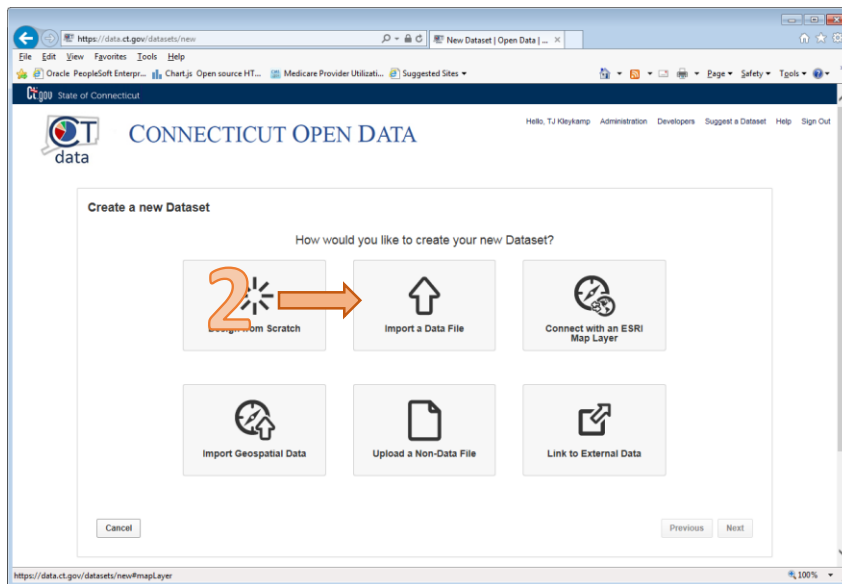
Once you have logged in to your Socrata Account, the following screen is where you will land. As a first time publisher, you won't see much here. However, as you continue to publish data or create maps and charts, you will see all of your content here.



Step 1.) Select “Create a new Dataset.”

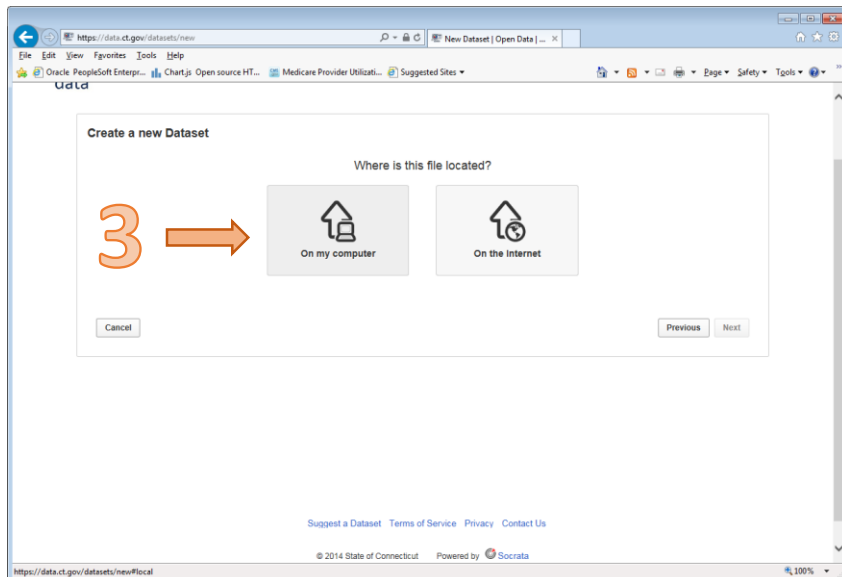
This will lead you to a new screen that provides a variety of options for loading different types of datasets or files in the Open Data Portal. The two most common types of files that are loaded are either Microsoft Excel spreadsheets (.xls or .xlsx files) or Geographic Information System files, generally known as “shapefiles” (.shp). First, we’ll upload an Excel file.

## Step 2.) Select “Import a Data File”

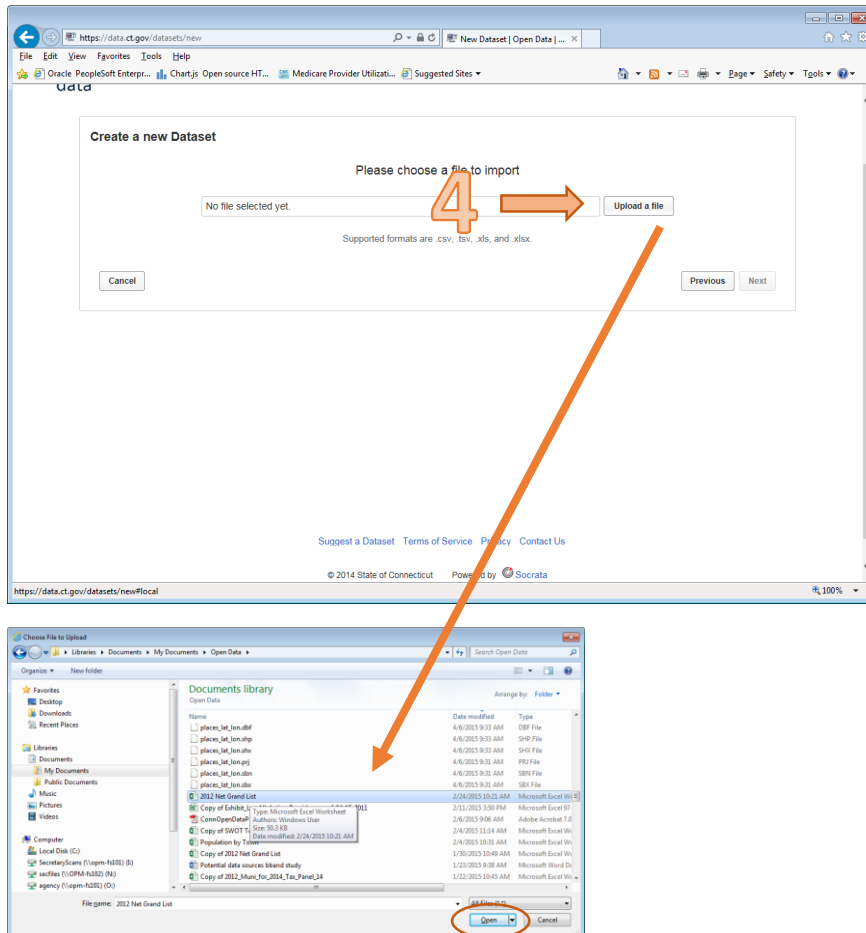


The next screen will ask you where your file is located. You can load files that are either on your computer, or on the internet. If the file is located on the internet, you'll need the URL for the file. For now, you can use one on your computer.

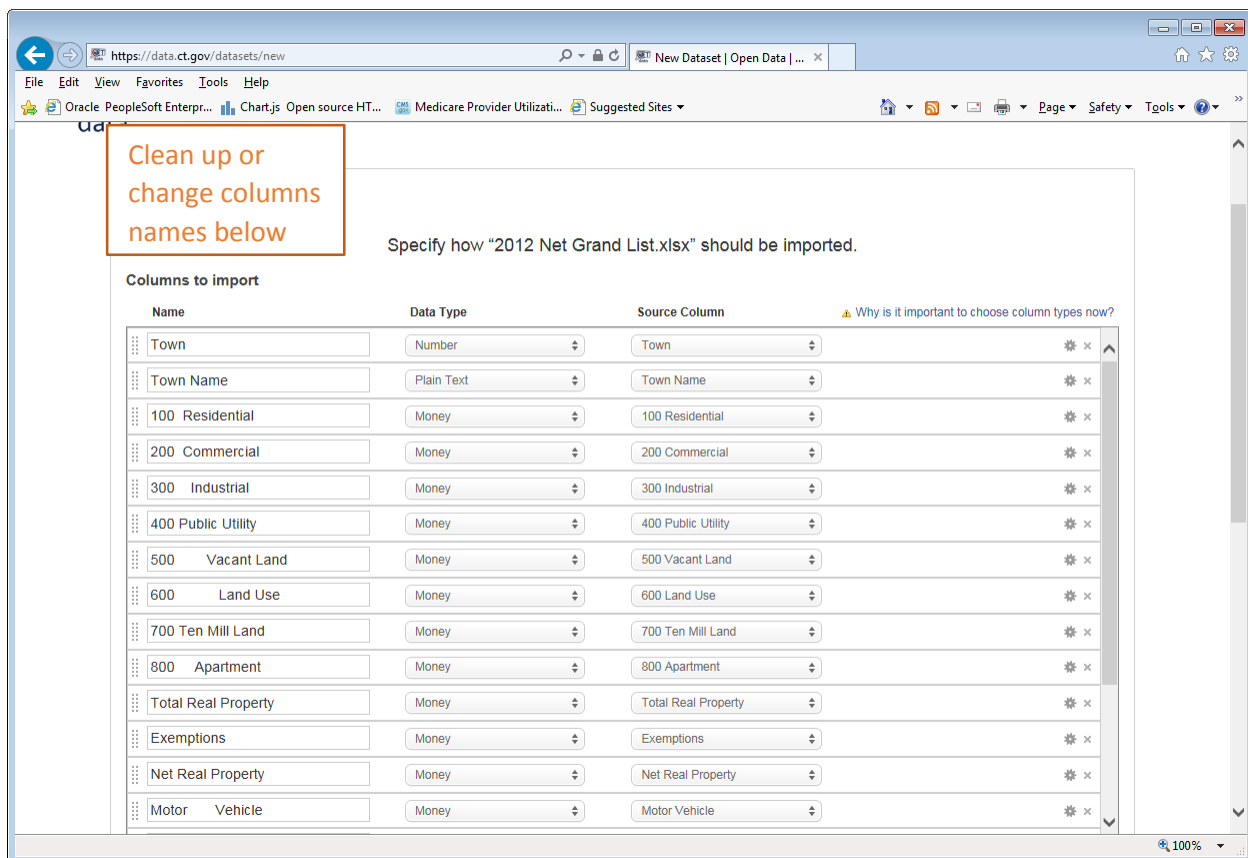
## Step 3.) Select “On My Computer”



Step 4.) In the next screen select “Upload a File.” This will launch Windows Explorer and allow you to navigate to the location of your file. Simply select the appropriate file and click “Open”



Once you have selected your file, the upload process will begin. Depending on the size, and number of columns, it may take a bit of time for the file to be processed. The Open Data Portal is analyzing your file to determine the column names, and the type of data that they contain (such as text, number, date, etc.). The following screen will allow you to make changes to your data, or the way that it is imported. You can modify the column names, data type, or even remove columns. For the most part, the Open Data Portal will recognize, and recommend the best way to import your data, but it's always good to verify that everything will be imported properly.



## LOCATION:

If your file has columns that reflect an address (address, street, city, state, zip code), the Open Data Portal will recognize those, and group them together to import as a "Location." The Open Data Portal has built in functionality to perform "geocoding" or "geolocation" on data sets, which is the process of taking a street address and assigning a latitude and longitude to it, so that the data can be placed on a map. This can be very useful, both to you as a publisher as well as users of the data. The Open Data Portal has a tool that will allow you to easily create interactive maps if you create a "Location" column.

However, when the dataset is uploaded this way, it groups all of these elements (address, city, state, zip code) into a single column which can impact the usefulness of your data. For instance, a user may no longer be able to sort or filter by City or Town. Fortunately, you don't have to choose between having a Location column and not having one. The Open Data Portal lets you do both! Therefore, it is recommended that you import data into a "Location" column, then add those additional elements as individual columns as follows:

Step 1.) Verify that the correct columns from your source spreadsheet are being imported to the appropriate "Location Source Columns." Sometimes, our spreadsheets do not contain a column for State, since often only deal with Connecticut locations. It's important then to enter the abbreviation "CT" as a custom value for State.



Specify how "SNAP\_RETAILERS.csv" should be imported.

**Columns to import**

Name	Data Type	Source Column
Store_Name	Plain Text	Store_Name
Address Line #2	Plain Text	Address Line #2
County	Plain Text	
Location 1	Location	

**Location Source Columns** Addresses will be geocoded as follows:

☒ Import from **multiple columns**

Existing Address Columns

Street: Address

City: ☒ City ☐ Enter a custom value

State: ☒ State ☐ Enter a custom value

Zip: ☒ Zip5 ☐ Enter a custom value

☐ Import from **single column**

Existing Latitude and Longitude Columns

Latitude: Latitude

Longitude: Longitude

Reset to preset: Suggested columns

**Headers**

Headers should be ignored and not imported as data. Please indicate how many rows of your source data you want to import.

Store_Name	Longitude	Latitude	Address	Address Line #2	City	State	Zip5	Zip4	County
S B MART Convenience Store	-72.878052	41.328117	411 Middletown Ave		New Haven	CT	6513	1007	NEW HAVEN
Rite Aid 3203	-72.698181	41.782837	1291 Albany Ave		Hartford	CT	6112	2198	HARTFORD
MR D'S MARKET	-73.092041	41.325996	78 Hawkins St		Derby	CT	6418	1531	NEW HAVEN
Collado Food Market	-72.695595	41.805309	655 Blue Hills Ave		Hartford	CT	6112	1208	HARTFORD

Step 2.) In order to add your Location columns as individual columns select the "Add New Column" button directly below the list of columns to import. A new empty column name will appear. A dropdown list will provide the available columns from your source spreadsheet. Simply select the appropriate value, and then type a name for it in the empty box.

Specify how "SNAP\_RETAILERS.csv" should be imported.

**Columns to import**

Name	Data Type	Source Column
Store_Name	Plain Text	Store_Name
Address Line #2	Plain Text	Address Line #2
County	Plain Text	County
Location 1	Location	

**Location Source Columns** Addresses will be geocoded after import is complete.

☒ Import from **multiple columns**

Existing Address Columns

Street: Address

City: ☒ City ☐ Enter a custom value

State: ☒ State ☐ Enter a custom value

Zip: ☒ Zip5 ☐ Enter a custom value

☐ Import from **single column**

Address: Number

Existing Latitude and Longitude Columns

Latitude: Latitude

Longitude: Longitude

Reset to previous suggested columns

**2** Give it a name

Select "Add New Column"

Select the appropriate source columns from your spreadsheet

Store Name	Longitude	Latitude	Address	Line #2	City	State	Zip5	Zip4	County
S B MART Convenience Store	-72.878052	41.328117	411 Mid				6513	1007	NEW HAVEN
Rite Aid 3203	-72.698181	41.782837	1291 Alb				6112	2198	HARTFORD
MR D'S MARKET	-73.092041	41.325996	78 Hawk				6418	1531	NEW HAVEN

Finalizing your import; The Open Data Portal will analyze your file and advise you if it seems that there may be errors in your upload. The example below shows that when the additional "Address" column was added, it was incorrectly set to import as a "Number" rather than text. In addition, it shows that the "Zip4" column from the source dataset was omitted from the upload. If there are no errors present, or you choose to ignore them (generally not advised), click the "Next" button to finalize your import.

Import from **single column**

Address Number Address

Reset to preset: Suggested columns Set Clear All + Add New Column

**Headers**  
 Headers should be ignored and not imported as data. Please indicate how many rows of your source file are headers:

Store_Name	Longitude	Latitude	Address	Address Line #2	City	State	Zip5	Zip4	County
S B MART Convenience Store	-72.878052	41.328117	411 Middletown Ave		New Haven	CT	6513	1007	NEW HAVEN
Rite Aid 3203	-72.698181	41.782837	1291 Albany Ave		Hartford	CT	6112	2198	HARTFORD
MR D'S MARKET	-73.092041	41.325996	78 Hawkins St		Derby	CT	6418	1531	NEW HAVEN
Collado Food Market	-72.695595	41.805309	655 Blue Hills Ave		Hartford	CT	6112	1208	HARTFORD

One of your rows is a header Fewer Rows More Rows

**Errors and Warnings**  
 For help, check out a list of [common warnings and errors](#).

- Column "Address" is set to import as **Number**, but our analysis shows that **Text** is a better fit. Should you choose to import as Number, roughly **100%** of your data will import incorrectly.
- Column "Zip4" is in your source data file, but is not currently set to be imported into your dataset.

Cancel Previous Next

[Suggest a Dataset](#) [Terms of Service](#) [Privacy](#) [Contact Us](#)

© 2014 State of Connecticut Powered by [Socrata](#)

https://data.ct.gov/datasets/new#

## METADATA:

Metadata is something that describes your dataset. In other words, it's data about data. It's also VERY important. Giving your dataset a meaningful title (without acronyms), a robust description that explains what it is, why it is collected, what it means, and what its limitations may be are all very important to potential users of the data. We typically provide this level of information on a web page, in a document, or with the spreadsheet when we distribute the data.

Below is an example of what you will likely see after your data is loaded:

**Create a new Dataset**

Please describe "SNAP RETAILERS"

\* Dataset Title: SNAP RETAILERS

Brief Description: Enter a description

Category: -- No category --

Tags / Keywords: Enter keywords  
 Enter one or more keywords separated by commas

Row Label: e.g. Crime, Person, Hospital  
 Describe what each row in the dataset represents (if applicable).

The following example provides a more descriptive title as well as additional description on the dataset itself. Your title and description should provide enough detail so that users can have a full understanding of what the dataset is and how it can be used. In addition, you should select a category that best fits your dataset. Note that these categories are fairly broad and intended to aid users who may be browsing the Open Data Catalog, and are not likely to know which agency published the data.

* Dataset Title	SNAP (Supplemental Nutrition Assistance Program) Authorized Retailers
Brief Description	A listing of retailers in Connecticut authorized to accept SNAP EBT benefits. The Supplemental Nutrition Assistance Program (SNAP), formerly known as Food Stamps, helps eligible individuals and families afford the cost of food at supermarkets.
Category	Health and Human Services
Tags / Keywords	snap, food, benefits, ebt Enter one or more keywords separated by commas
Row Label	e.g. Crime, Person, Hospital Describe what each row in the dataset represents (if applicable).

Also, note above the use of “Tags/Keywords” which can aid in searches.

Below is the list of Categories available, and how they are reflected on the homepage

Category	<div style="border: 1px solid black; padding: 2px;"> -- No category --  Business  Education  Environment and Natural Resources  Government  Health and Human Services  Housing and Development  Public Safety  Transportation </div>
Tags / Keywords	
Row Label	

**Homepage**

Categories

- [Browse All](#)
- [Business](#)
- [Government](#)
- [Education](#)
- [Environment & Natural Resources](#)
- [Health & Human Services](#)
- [Housing & Development](#)
- [Public Safety](#)
- [Transportation](#)

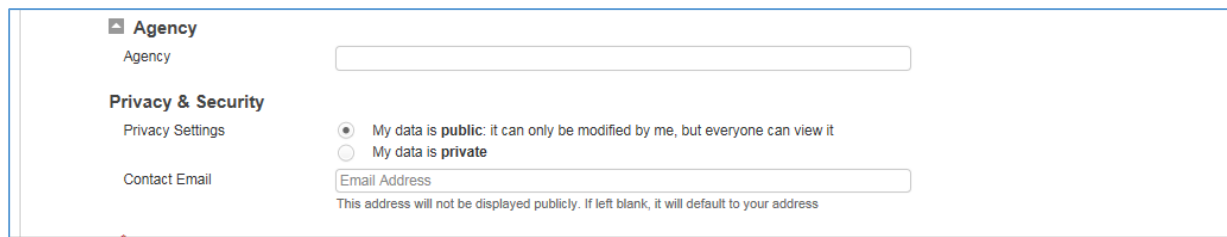
**License Type:** In general, CT government data is “Public Domain” however, because the software for the Open Data Portal is used by governments around the world, other options are available. It is important to explicitly identify your data as “Public Domain” so that users know there are no legal restrictions associated with the use of the data. A “Creative Commons” license may also be used, which simply requires that users of your data attribute the source in any publications or projects in which it is used.

<b>Licensing &amp; Attribution</b>	
License Type	<div style="border: 1px solid black; padding: 2px;"> -- No License --  Public Domain  Italian Open Data License 2.0  UK Open Government Licence v3  Creative Commons  Open Data Commons </div>
Data Provided By	
Source Link	
<p><b>Attachments</b></p> <p>To upload attachments, first finish publishing this dataset. Then, click <a href="#">About</a> in the toolbar, and click Edit Metadata. You will then be able to upload attachments.</p>	
<p><b>Microsites</b></p>	

**Data Provided By:** This should be used to specify the State Agency providing the data (e.g Office of Policy and Management) – Do not use an acronym (OPM)

**Source Link:** If there is a web page that provides additional information about the dataset, agency, or program the url (website address) it should be included here

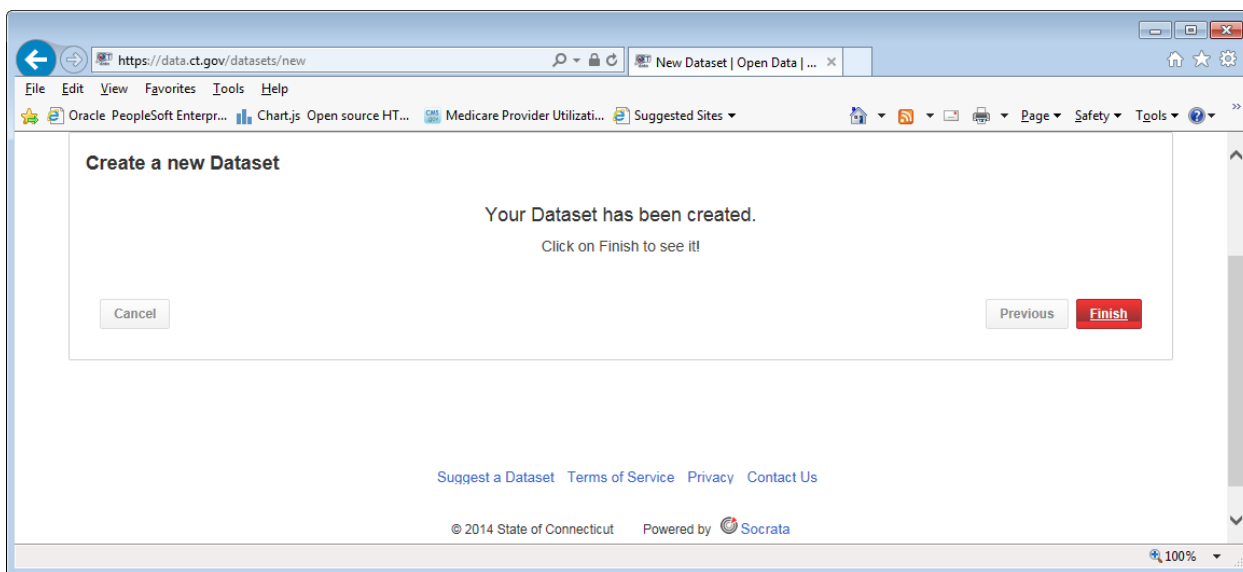
**Agency:** This area is reserved for future use, and will include a dropdown list of State Agencies



The screenshot shows a form with two main sections. The first section is titled 'Agency' and contains a single text input field. The second section is titled 'Privacy & Security' and contains two radio button options: 'My data is public: it can only be modified by me, but everyone can view it' (which is selected) and 'My data is private'. Below these options is a text input field labeled 'Email Address' with a small note underneath stating 'This address will not be displayed publicly. If left blank, it will default to your address'.

**Privacy Settings:** Generally you should select “Public” as the intent of the Open Data Portal is to provide public access to data. However, the Portal can be used to privately share data between 2 or more users. However, the Open Data Portal should never be used to share any data which could contain Personally Identifiable Information (PII) or Protected Health Information (PHI) that is protected by state or federal law.

Finally, click Finish. However, your data won’t be publicly visible just yet.



The screenshot shows a web browser window with the URL 'https://data.ct.gov/datasets/new'. The page title is 'New Dataset | Open Data | ...'. The main content area has the heading 'Create a new Dataset' and a message that says 'Your Dataset has been created. Click on Finish to see it!'. There are three buttons: 'Cancel', 'Previous', and 'Finish' (which is highlighted in red). At the bottom of the page, there are links for 'Suggest a Dataset', 'Terms of Service', 'Privacy', and 'Contact Us'. The footer includes the text '© 2014 State of Connecticut' and 'Powered by Socrata'.

You now should have a “Working Copy” of your data set. For now, it can only be seen by you and the Site Administrator. The purpose is to allow you to inspect your data to make sure it imported properly, and to make any final formatting changes. If everything looks good, click “Publish Dataset”

https://data.ct.gov/dataset/DEMO-SNAP-RETAILERS/pf3h-2vtz DEMO-SNAP RETAILERS | ...

File Edit View Favorites Tools Help

Oracle PeopleSoft Enterpr... Chart.js Open source HT... Medicare Provider Utilizati...

State of Connecticut

CT data

CONNECTICUT OPEN DATA Hello, TJ Kleykamp Administration Developers Suggest a Dataset Help Sign Out

[Working Copy] DEMO-SNAP RETAILERS

Publish Dataset

DO NOT USE - FOR DEMONSTRATION ONLY

Find in this Dataset

Edit Manage Filter Export About

	Store_Name	Address Line #2	County	Location 1	Address
1	S B MART Convenience Store		NEW HAVEN	411 Middletown Ave	
2	Rite Aid 3203		HARTFORD	1291 Albany Ave	
3	MR D'S MARKET		NEW HAVEN	78 Hawkins St	
4	Collado Food Market		HARTFORD	655 Blue Hills Ave	

Finally, your dataset is now "Open!"

https://data.ct.gov/dataset/DEMO-SNAP-RETAILERS/pf3h-2vtz DEMO-SNAP RETAILERS | ...

File Edit View Favorites Tools Help

Oracle PeopleSoft Enterpr... Chart.js Open source HT... Medicare Provider Utilizati...

State of Connecticut

CT data

CONNECTICUT OPEN DATA Hello, TJ Kleykamp Administration Developers Suggest a Dataset Help Sign Out

DEMO-SNAP RETAILERS

DO NOT USE - FOR DEMONSTRATION ONLY

Find in this Dataset

Edit Manage More Views Filter Visualize Export Discuss Embed About

	Store_Name	Address Line #2	County	Location 1	Address
1	S B MART Convenience Store		NEW HAVEN	411 Middletown Ave	
2	Rite Aid 3203		HARTFORD	1291 Albany Ave	
3	MR D'S MARKET		NEW HAVEN	78 Hawkins St	
4	Collado Food Market		HARTFORD	655 Blue Hills Ave	
5	Clean and Friendly		NEW HAVEN	1351 North Main Street	
6	PRICE CHOPPER 205	Tri City Shopping Center	TOLLAND	35 Talcottville Rd	
7	7 - Eleven Food Store 19960		HARTFORD	276 S Main St	
8	Brook Market		HARTFORD	139 Brook St	
9	TURNPIKE BP, Inc.		NEW HAVEN	750 1st Ave	
10	Three Corners Grocery		NEW HAVEN	453 Huntington St	
11	Deangely Market		NEW HAVEN	362 Oakville Ave	
12	Sams Food Stores 1062		HARTFORD	1149 New Britain Ave	
13	WALMART 5095		HARTFORD	495 Flatbush Ave	

Suggest a Dataset Terms of Service Privacy Contact Us

© 2014 State of Connecticut Powered by Socrata

100%


https://data.ct.gov/dataset/DEMO-SNAP-RETAILERS/pf3h-2vtz

DEMO-SNAP RETAILERS | ...

FileEditViewFavoritesToolsHelp

Oracle PeopleSoft Enterpr...Chart.js Open source HT...Medicare Provider Utilizati...

State of Connecticut



CONNECTICUT OPEN DATA

Hello, TJ KleykampAdministrationDevelopersSuggest a DatasetHelpSign Out

DEMO-SNAP RETAILERS


DO NOT USE - FOR DEMONSTRATION ONLY

EditManageMore ViewsFilterVisualizeExportDiscussEmbedAbout

Find in this Dataset

	Store_Name	Address Line #2	County
1	S B MART Convenience Store		NEW HAVEN
2	Rite Aid 3203		HARTFORD
3	MR D'S MARKET		NEW HAVEN
4	Collado Food Market		HARTFORD
5	Clean and Friendly		NEW HAVEN
6	PRICE CHOPPER 205	Tri City Shopping Center	TOLLAND
7	7 - Eleven Food Store 19960		HARTFORD
8	Brook Market		HARTFORD
9	TURNPIKE BP, Inc.		NEW HAVEN
10	Three Corners Grocery		NEW HAVEN
11	Deangely Market		NEW HAVEN
12	Sams Food Stores 1062		HARTFORD
13	WALMART 5095		HARTFORD

About This Dataset



**TJ Kleykamp**  
created Aug 19, 2015  
updated Aug 19, 2015

Edit Metadata

View Full Page

Description


DO NOT USE - FOR DEMONSTRATION ONLY

Activity

Community Rating

Your Rating

Suggest a DatasetTerms of ServicePrivacyContact Us

© 2014 State of ConnecticutPowered by  Socrata

https://data.ct.gov/dataset/DEMO-SNAP-RETAILERS/pf3h-2vtz/edit\_metadata100%